

# Weekly Report (2018.12.3-2018.12.9)

## DONE

### 1. size 感知投稿项目：

#### • 12.3

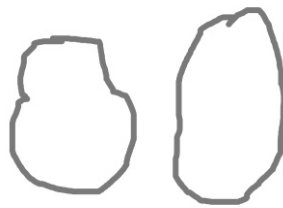
1. 继续安排做实验
2. 把近期所有结果扔进结果分析系统跑了一下，总结如下：
  - density 的结果最好，方差普遍很小，scale-bias 关系图呈现明显的线性，并且与 radius-bias 关系图的结果对应的很好。
  - correlation 继上次出现线性关系后，有一次出现了线性，说明上次的结果不是偶然，打破了之前对 correlation 下的判决，尺寸对 correlation 没有明显影响。
  - cluster separation 是做的最不好的，有的人不管怎么训练，方差一直都很大，所以在这一大组实验中，有不少人方差处于2和3之间。如果删除结果的话，要删掉不少。

#### • 12.4

1. 与被试进行沟通，对 cluster 实验进行微调与测试。微调如下：

通过与几个做的不好的被试进行沟通，发现一个在判断的时候的一个共性的问题，看不出两个聚类，不知道中间重叠的点属于哪一个聚类，所以通过轮廓以及散点区域较长方向的长度来判断。但是，两个 cluster 的散点区域虽然都是圆形，但不是完美的圆（尤其高斯分布的时候，圆形不会撑得很饱满，均匀分布相对好一些），这就使得看轮廓分辨聚类分离程度会差很多，而恰恰有很多人倾向于看轮廓来分辨区分两个聚类。

如下图所示的一个典型的散点图的轮廓：



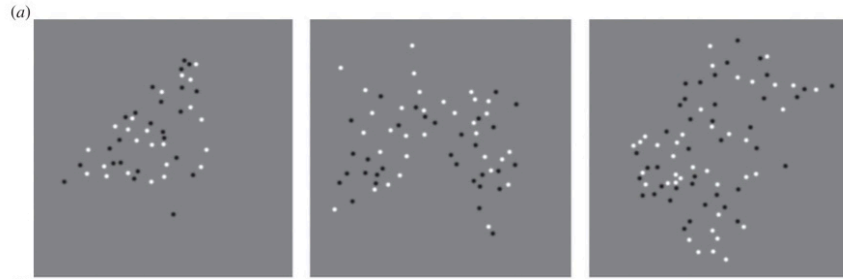
这两个散点图的实际情况是：左边的聚类分离程度小于右边的。

在选择哪边的聚类分离程度更大时，多数人倾向于选择左边的图，左边的图的轮廓更能让人觉着这是两个cluster，并且能分开，而右边的散点图没有明显的轮廓，让人觉着他没有分开。在尺寸相等的时候，可能会因为右边比左边的图长而选择右边，但是尺寸不同时，很大可能选择左边。还存在一个情况就是，被试在经过几次训练以后，会凭经验，选择正确的右边，并不是出于直觉。

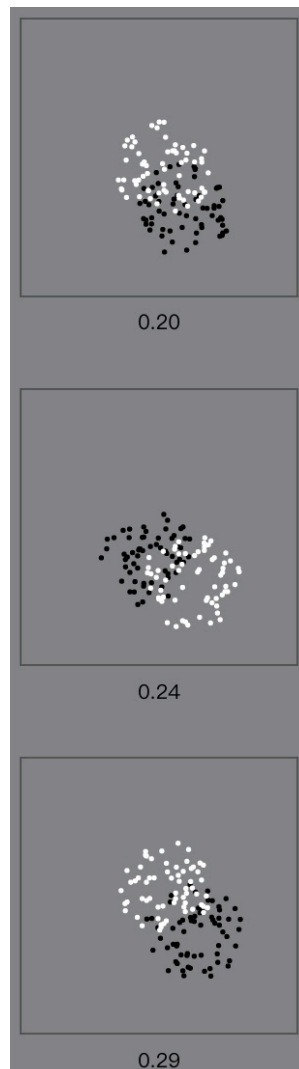
基于上面的反馈，作出以下微调：

- 将两个聚类的散点通过颜色区分开。
  - 已有的研究聚类的工作均通过不同的颜色将不同聚类区分开。

- 在实际应用中，我们看聚类，都是通过编码区分开的聚类，而我们研究的是尺寸变化对聚类分离程度的感知偏差，首先我们还是要能看出两个聚类，然后判断分离程度的大小。（不知道这么理解是否正确？）
- 对于颜色的选取，如果使用彩色，颜色选取又很讲究。之前查到的一些相关工作，他们使用的是深灰色背景，黑白两种颜色的点，如下图所示，有相当多工作使用这种配色，其中包括研究密度与数量的文章，但是都没有在文章中注明原因。研究 cluster 的工作多使用彩色。在我们的实验中，只设置了两个聚类，我觉得黑白两种颜色刚刚好。



参考文献图



实验用图

- 为生成的数据添加聚类的标签，以便在绘制时使用不同的颜色

微调后的实验反馈：

- 比之前的实验简单许多，可以通过中间相交的部分来辅助判断。
- 深灰色背景，有一种夜间模式的感觉，眼睛没那么容易疲劳。

- 找之前做的方差比较大的被试重新做做的不好的那部分实验，做出来的结果有很大改善。由于修改以后的实验做的还比较少，从整体结果上还没看出明显改善。

## 2. 继续实验。

- 12.5

1. 集中找了几个只做人只做了新版的 cluster 实验，看了下结果，除了实验1的高斯分布的 scale-bias 关系还不是很漂亮，其余的三个大致有线性出来了，而且单个人的方差基本在2以下。明天再多做些结果。

- 12.6-12.7

1. 安排做实验，实验室能做的人差不多已经都做了，在微信群以及论坛里重新发帖，一个人不需要完整做完所有实验，可以只做其中1-2大组的实验，约了一拨人。

## 2. 时间片截取项目调研

1. 与王杰讨论旧版代码实现，探讨 TCPTree 中时间片切分的策略。

## 3. 组会报告

Charticulator: Interactive Construction of Bespoke Chart Layouts ( InfoVis 2018 )

这篇文章提出一种交互式创作工具Charticulator，可以创建定制和可重复使用的图表布局。

## 小结

工作日工作时长11+10.5+7.5+9+7，周末4小时，总时长约49h。由于 size 感知实验的实验总时长有点长，找不到做实验的人，所以对实验策略做了调整，一个人不需要完整做完所有实验，可以只做其中1-2大组的实验，修改策略后，报名做实验的人多起来了。

## PLAN

### 短期计划（一周）

1. size 感知投稿项目：约人做实验，看结果。
2. 着手准备动态图中时间截取工作。
3. 组会报告博客。

### 中期计划

1. Visevo论文
2. 动态图时间片截取项目：目前进度有些慢了，要注意在做 size 感知项目的同时找时间进行调研并做些实验。
3. 综述

### 长期计划

1. 学习更多机器学习、数据挖掘相关的算法。
2. 在项目中锻炼自己的思考能力与代码能力。